

Ethics in classroom assessment practices: Issues and attitudes

Susan K. Green^{a,*}, Robert L. Johnson^b, Do-Hong Kim^c, Nakia S. Pope^a

^aWinthrop University, Rock Hill, SC 29733, USA

^bUniversity of South Carolina, SC, USA

^cUniversity of North Carolina at Charlotte

Received 24 January 2006; received in revised form 5 April 2006; accepted 10 April 2006

Abstract

Student evaluations should “be ethical, fair, useful, feasible, and accurate” [JCSEE (2003). *The student evaluation standards*. Arlen Gullickson, Chair. Thousand Oaks, CA: Corwin]. This study focuses on defining ethical behavior and examining educators’ ethical judgments in relation to assessment. It describes the results from a web-based survey of educators in which they read a brief scenario and indicated whether the student evaluation practice in the depiction was ethical or unethical. Results showed strong agreement among the educators on fewer than half of the scenarios presented in this study. These findings suggest that assessment is currently an educational realm without professional consensus.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Evaluation methods; Ethics; Classroom assessment

1. Introduction

The importance of the evaluation of students is evidenced in the recent publication of *The Student Evaluation Standards* (The Joint Committee on Standards for Educational Evaluation [JCSEE, 2003]). The standards are offered as “principles that should guide and govern student evaluations” (JCSEE, p. xx) and “require that student evaluations be ethical, fair, useful, feasible, and accurate” (JCSEE, p. 3).

The need for guidelines on ethical assessment practices is evident in frequent incidents taken from newspaper headlines. In one incident, a biology teacher in the Midwest USA decided to assign students failing grades for the science course after

the students were caught cheating on a class project. The decision split the community and the teacher resigned (Carroll, 2002). In another instance, the state law enforcement division in a southern USA community investigated a teacher after she used in her classroom some commercially available test preparation materials that are designed to simulate the type of items and subject matter in the state test. Finally, the president of a southeastern USA college recently fired two professors for their refusal to adhere to a policy awarding freshman 60% of their grade based on effort (Click, 2004). The dismissal of the professors brought unfavorable national attention to the administrator’s misguided policy on effort and grades.

Guidelines for assessment practices include those offered in *The Student Evaluation Standards* (JCSEE, 2003), the *Standards for Teacher Competence in the Educational Assessment of Students*

*Corresponding author.

E-mail address: greens@winthrop.edu (S.K. Green).

(American Federation of Teachers, National Council on Measurement in Education, & National Education Association, 1990), the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), and the Code of Ethical Conduct and Statement of Commitment of the National Association for the Education of Young Children (2005). As Gipps (1994) points out, however, guidelines are often “general and almost superficial” (p. 144). They provide rules without an overarching ethical framework that can provide support and guidance for exercising judgment in specific situations.

Several authors have provided ethical guidelines related to preparation for high stakes testing (e.g., Haladyna, Nolen, & Haas, 1991; Mehrens & Kaminski, 1989; Popham, 1991). In addition, several studies have examined in-service and pre-service teacher attitudes regarding the appropriateness of standardized test preparation strategies (e.g., Bright, 1992; Kher-Durlabhju & Lacina-Gifford, 1992; Moore, 1993). However, ethical standards related to testing are not consistently defined or agreed upon (Thorndike, Cunnningham, Thorndike, & Hagen, 1991; and see Kilian [1992] in response to Popham [1991]). Gipps and Murphy (1994) go a step farther and suggest that more attention should be focused instead on wider equity issues such as equality of access to instruction. The lack of agreement highlights the need for an overarching ethical framework from which to develop the capacity to make judgments about ethical assessment practices.

Other authors have discussed the importance of ethics in assessment (e.g., Baumgart, 1996; Gipps, 1994; Popham, 2000). For example, Baumgart (1996) applauds the development of ethical codes related to assessment. He argues that education, reflection, and self-regulation related to ethics are more likely with easy access to such codes of ethics, but does not suggest principles that might govern such activities. Gipps (1994) points out that the powerful impact of assessment necessitates an ethical framework on which to draw, but limits the discussion to specific issues such as consequential validity (Messick, 1993), equity, and test preparation. Popham (2000) does offer an ethical guideline—“when teachers engage in test preparation practices which, if brought to the public’s attention, would discredit the education profession,

such practices may be considered professionally unethical (p. 82).” While Popham does offer this guideline, he neglects to place it within a larger ethical framework. Why, for example, does the ethicality of test preparation depend on what other people think? Making ethical assessment practices contingent on whether or not one may get caught results in a simplistic utilitarian calculus, which is not, in and of itself, ethical.

Some textbooks do briefly discuss ethical principles that can guide ethical judgments related to assessment. For example, Airasian (2005) suggests that the ethical standards for assessment refer to “some aspect of a teacher’s fairness in dealing with his or her pupils (p. 20).” Similarly, Taylor and Nolen (2005) point out that because poor assessment can significantly affect students, “the ethical responsibility of educators is ‘first, Do No Harm’ (p. 7).” Payne (2003) sums up ethical codes by stating that educators should “Assess As Ye Would Be Assessed.” These principles are not elaborated as tools to assist educators in making day-to-day judgments, however. To summarize, the discussion in the literature often describes what various authors consider ethical practices, but these discussions suffer from a lack of explicit identification of the underlying ethical assumptions related to making decisions about behavior in the realm of assessment.

The present authors define ethical behavior as acting based on one’s judgment of an obligation—a duty by virtue of a relationship with a person, persons, or social institution (Kant, 1785; Rawls, 1971). What makes a behavior ethical or unethical is whether that behavior is consistent with or contradictory to one’s obligations. Given these conditions, there is always a gap between principles and behavior requiring judgment in a specific situation (Hostetler, 1997; Kant, 1790; Strike & Soltis, 1998), so uniform agreement on standards in every case would not be expected. Educators must be given the space, autonomy, and support to learn to use their judgment. This approach thus shows an important distinction between ethics and best practices in teaching and assessment. Ethics is addressed in terms of obligations based on relationships so there is an affective component as well as a rational one (Noddings, 1984; Nussbaum, 1986, 1990). More specifically, obligations are felt, but best practices are merely followed.

Other literature suggests that teachers may not always be well equipped to make ethical judgments

related to assessment. Teachers' knowledge base about guidelines or ethical codes may be uncertain because they often lack formal assessment training or their training has become dated (Impara, Plake, & Fager, 1993; Plake & Impara, 1997; Stiggins, 1999). Conflicting norms related to teaching (e.g., institutional norms vs. ethical interpersonal norms) often place teachers in ethical dilemmas (Colnerud, 1997). But, teachers' ability to apply guidelines to a *specific* context has received limited attention in the literature. One exception is the Plake and Impara (1997) study in which five items in a test of teacher assessment competence measured teachers' ability to answer items related to ethical issues in student evaluation. These items, however, all focused on issues relating to standardized tests.

Another element that has not been thoroughly addressed in the literature is the application of ethical principles to ongoing classroom assessment. In the classroom, teachers are required to weigh their obligations and make judgments every day related to assessment. Comprehensive principles related specifically to classroom evaluation have not been widely promulgated, although several assessment texts do discuss ethical codes and practices related to classroom assessment (e.g., Airasian, 2005; Taylor & Nolen, 2005).

As a starting point for generating principles widely applicable to the various types of assessment that occur in classrooms, the present authors examined professional standards, articles addressing ethical principles in education, papers related to ethical standardized test preparation practices, and assessment texts. Two general guiding principles seem to capture the essence of the ethical concerns in these documents.

The first is to *Do No Harm* (Taylor & Nolen, 2005). This is a basic broad ethical principle that people in general use to govern their lives by virtue of their being human. It is a variant of ethical principles handed down through the centuries, like the Golden Rule (or Assess as Ye Would Be Assessed [Payne, 2003]). (This is a core ethical principle in other professions such as medicine [Jonsen, Siegler, & Winslade, 1998; Munson, 2000].) The judgment involved comes in defining what harm is, or choosing between different harms in the relationship between teacher and student. This principle stems from the basic premise that ethical guidelines must protect the rights of individuals affected by an evaluation. This premise is congruent with the proprietary standards in the JCSEE (2003)

Student Evaluation Standards (p. 7). Several of those standards address aspects of this fundamental assumption: to serve the needs of students, to treat students with respect, and to incorporate basic principles of fairness (Airasian, 2005). Popham's (1991) point that educators serve *in loco parentis* also lends support to the principle that assessment should protect students and Do No Harm to them.

In the present conceptualization, the guiding principle *Do No Harm* was chosen to make the implications more concrete, focusing on the harm that can be done when the basic premise is *not* followed. Fairness (or protection of student rights) is a general principle that no one contests in the abstract. However, thinking about causing harm focuses the discussion at the level of the implications of everyday practice. Educators must be well versed in the potential impact of the practices they use because their assessment and evaluation may have a variety of unintended consequences for their students. For example, a teacher who uses surprise items on a test that did not appear on the study guide may do harm by breaking the implicit bond of trust between teacher and student. A teacher who passes out tests from highest grade to lowest may do harm by breaching confidentiality. Such actions imply lack of respect for student rights and needs.

The second general guiding principle for classroom assessment is to *Avoid Score Pollution*. This is a specific iteration of the *Do No Harm* principle that applies to assessment. This principle is adapted from guidelines suggested by both Popham (1991) and Haladyna et al. (1991) for ethical standardized test preparation. Both authors suggest that any practice that improves test performance without concurrently increasing actual mastery of the content tested produces score pollution. That is, the score on the test does not represent actual student achievement in the content area and is "polluted" by factors unrelated to academic attainment. If scores do not reflect mastery then harm has been done. This situation is akin to lying. For example, practicing beforehand with actual test content would produce score pollution. In essence, this is a validity issue. Test scores no longer measure generalized mastery but simply ability to memorize specific test items.

The present authors believe this principle can be extended to other elements of classroom assessment. Any official assessment, including grades, should reflect only the extent to which students have mastered the goals of instruction (e.g., Brookhart,

2004; Smith, Smith, & DeLisi, 2001). When educators modify grades or scores because of student effort, late work, or behavior problems, for example, the scores do not accurately communicate the level of mastery. Similarly, many teachers do not use a blind grading system and may unconsciously prefer certain students. These educators may unintentionally engage in score pollution by giving less favored students lower grades than they deserve. Thus classroom grades are “polluted” by these other factors. Students, their families, and other stakeholders in the education system need valid information regarding academic achievement.

1.1. Conceptual framework and research questions

The present study draws on both theoretical and empirical foundations in the areas of ethics and assessment. From the ethics literature derives the assumption that ethical behavior is defined as acting based on one’s judgment of an obligation. In addition, the notion of an ever-present gap between principles and ethical behavior in specific circumstances requires a focus on specific judgments as well as on general principles such as *Do No Harm*. From the assessment literature comes the assumption that assessment of student achievement should accurately reflect mastery of content, should be administered fairly, and should be treated confidentially. Resulting empirical studies in this tradition have focused primarily on teachers’ knowledge base about practices related to standardized testing. In the present study, we wished to explore the intersection of these two bodies of literature by examining specific circumstances requiring ethical judgments about a variety of classroom assessment practices.

The guiding research question was, “What is the degree of agreement among educators about specific ethical questions related to assessment issues that arise in the classroom?” Evidence of strong agreement would suggest that educators use similar guidelines and methods of judgment in enacting assessment in the classroom. Evidence of weak agreement would suggest that further exploration and dialogue related to ethical principles and their application in specific circumstances would be useful. In particular, clarification of how the process between principle and judgment takes place would be needed. Thus, the present study was designed to survey educators by presenting a variety of class-

room assessment scenarios to determine their ethical position on a range of assessment practices.

2. Method

2.1. Participants

The participant pool for this study consisted of educators in the graduate and undergraduate programs at two major southern USA universities. All of these individuals were enrolled in a classroom assessment course or a course with a classroom assessment component. Participants completed the survey prior to the class session that addresses ethics in classroom assessment. The pool included pre-service teachers, in-service teachers, and administrators. This study focuses on the responses of pre-service and in-service teachers.

2.2. Instrument

The current study organizes the guidelines for ethical student evaluation into a framework that addresses both classroom assessment and standardized testing. Thus, scenarios for items on both classroom assessment and classroom issues related to standardized testing were developed using resources from the assessment literature (e.g., *The Student Evaluation Standards* [JCSEE, 2003] and *The Principles for Fair Student Assessment Practices for Education in Canada* [Joint Advisory Committee, 1993]) and the authors’ experiences as well as anecdotes gathered from graduate students over several semesters.

The researchers used a web-based survey to present a series of scenarios in which an educator is engaging in a student evaluation practice that may or may not be ethical. For example, one scenario states that “A high school social studies teacher bases students’ final semester grade on 2 multiple choice tests.” In completing the survey item, the educator selects the *Ethical* or *Unethical* option.

2.3. Procedure

The first draft of the survey consisted of 36 scenarios and 6 questions about demographic information. A pilot survey was conducted with 74 participants in the spring of 2004. Participants were given oral instructions during class on how to access the survey web site and complete the web-based

survey. The results of the pilot survey were reviewed and six items that appeared confusing were modified or replaced.

The final 36-item survey was administered in the summer and fall of 2004. Of 169 responses, 114 were pre-service and 55 were in-service teachers.

3. Results

3.1. Analyzing agreement and disagreement

Seven categories were created based on the topics of the scenarios: Standardized Test Preparation, Standardized Test Administration, Multiple Assessment Opportunities, Communication about Grading, Grading Practices, Bias, and Confidentiality. For the total group of 169 respondents, the categories with highest agreement on individual items among respondents were calculated by determining the percentage of items in that category on which at least 80% of respondents agreed. For example, 90% of educators indicated it is ethical to for a teacher to spend a class period to train his students in test-taking skills (e.g., not spending too much time on one problem, eliminating impossible answers, guessing) prior to testing. Similarly, 96% of educators indicated it was unethical when a teacher provides a hint about a child's incorrect answer during the administration of a standardized test. In both instances, educators were in agreement in their views about whether a practice was ethical or unethical. We selected the 80% level because it indicates considerable agreement among respondents. In addition, few items were selected by 90% of respondents as being ethical or unethical; so little would be gained by such an analysis.

We also examined the areas of disagreement. When 50% of teachers indicate a practice is ethical and 50% label the practice unethical, then the split (i.e., disagreement) can be no greater. For example, high disagreement occurs when 57% of teachers indicate that the weighting of homework heavily in determining report card grades is ethical, and 43% indicate such a practice is unethical. Teachers, thus, are split on this issue. Items with percentages from 50% to 70% were identified as displaying substantial disagreement among teachers.

3.2. Analysis within content categories

The level of agreement for the individual items in *Standardized Test Preparation* ranged from 53% to

90% (see Table 1). Items that fell within 80% agreement included a teacher spending time training students in test-taking skills (Item 5), a teacher administering a parallel form of a norm-referenced state test (Item 6), and a teacher using a commercially available publication with the same format and skills as a state test (Item 12). One item reflected high levels of disagreement (falling within the 50–70% range): “Based on his review of the district's mathematics frameworks, a teacher creates learning activities with specific math problems that are included in the annual achievement test (Item 4).”

In the category *Test Administration*, as shown in Table 2, almost all (96%) of the in-service and pre-service teachers rated as unethical a teacher drawing a student's attention during testing to a problem that the child had missed (Item 7). The other item in this category, in which a teacher corrects a student who is marking items out of sequence, fell into the high disagreement range, with 69% indicating that the practice is ethical and 31% indicating unethical.

The rates of agreement for *Multiple Assessment Opportunities* (see Table 3) ranged from 78% to 99%. Most recognized as ethical the use of many forms of assessment in the classroom (99%, Item 21). A large percentage rated as unethical the questionable practice of relying on one form of assessment in a classroom (85%, Item 25). Reliance on too few assessments for determining grades (78%, Item 10) nearly reached the 80% agreement level.

As seen in Table 4, for the category of *Communication about Grading*, 98% of in-service and pre-service teachers rated as ethical the appropriateness of providing students with a written policy about the calculation of grades (Item 31), informing students about the materials that are important in preparing for a test (Item 27), and stating the manner in which a task will be graded when it is assigned (Item 1). The other item in this category, using surprise items not on the study guide, fell into the high disagreement range, with 34% of both groups rating it as ethical and 66% as unethical.

As shown in Table 5, in-service and pre-service teachers' views about *Grading Practices* varied widely. Levels of agreement ranged from 57% to 86%. Items where over 80% agreement was reached included rating as ethical a teacher basing grades for a group project on the group product plus a heavily weighted individual component (Item 3, 83%), rating as unethical a teacher lowering report card

Table 1

Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices in standardized test preparation

Item #	Scenarios for standardized test preparation	Respondents' answers	Total (N = 169), %	Pre-service teachers (N = 114), %	In-service teachers (N = 55), %
5	A teacher spends a class period to train his students in test-taking skills (e.g., not spending too much time on one problem, eliminating impossible answers, guessing).	Ethical Unethical	89.9 10.1	89.5 10.5	90.9 9.1
12	A teacher uses scoring high on the MAT, a commercially available publication with the same format and skills as the Metropolitan Achievement Test (but not the same items), in preparation for state testing.	Ethical Unethical	88.8 11.2	86.6 11.4	89.1 10.9
6	A teacher administers a parallel form of a norm-referenced achievement test to her students in preparation for the state testing. The parallel form is another version of the state test that assesses the same content; however, the items on the parallel form are not the same ones as on the state form of the achievement test.	Ethical Unethical	80.5 19.5	79.8 20.2	81.8 18.2
2	A teacher adds vocabulary words from a standardized, norm-referenced verbal aptitude test to classroom vocabulary tests.	Ethical Unethical	74.0 26.0	81.6 18.4	58.2 41.8
23	An elementary teacher quizzes students in the lunch line about the number of pints in a quart because students had missed the item on previous administrations of the state standardized test.	Ethical Unethical	70.4 29.6	68.4 31.6	74.5 25.5
4	Based on his review of the district's mathematics frameworks, a teacher creates learning activities with specific math problems that are included in the annual achievement test.	Ethical Unethical	53.3 46.7	63.2 36.8	32.7 67.3

grades for disruptive behavior (Item 35, 85%), rating as ethical a teacher considering student effort when determining grades (Item 30, 85%), and rating as ethical lowering grades for late work (Item 24, 86%).

Grading Practices items with a high level of disagreement included a physical education teacher giving a student a zero as a homework grade for not returning a form requiring a parent's signature (Item 8, 57% unethical), deducting more points on a test for a wrong answer than leaving an item blank (Item 14, 69% unethical), changing a grade from a B+ to an A because a student had mastered course objectives but not completed all homework assignments (Item 13, 63% unethical), and weighting homework heavily in determining report card grades (Item 32, 57% ethical).

Agreement for *Bias Issues* (Table 6) ranged from 52% to 94%. Most respondents (94%) agreed to the ethicality of a teacher allowing a student with a learning disability to use accommodations in testing (Item 34). They also agreed that it was unethical to restrict the number of As in a class based on a belief that student work is rarely perfect (Item 22, 80% agreement).

Two *Bias* items fell in the strong disagreement range in this category. First, 59% of all respondents rated as unethical the practice of only addressing student strengths in narrative report cards (Item 36). Second, 52% rated as unethical the practice of grading essay tests with the knowledge of student identities (Item 26).

In the *Confidentiality* scenarios, agreement was high, ranging between 76% and 98% (see Table 7).

Table 2
Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices in standardized test administration

Item #	Scenarios for standardized test administration	Respondents' answers	Total (N = 169), %	Pre-service teachers (N = 114), %	In-service teachers (N = 55), %
17	While administering a standardized test, a teacher notices that a child has skipped a problem and is now recording all his answers out of sequence on the answer form. The teacher stops at the child's desk and shows the student where to record the answer he is working on and instructs him to put the answer to each question with the same number on the answer sheet.	Ethical Unethical	69.2 30.8	68.4 31.6	70.9 29.1
7	While administering a standardized test, a teacher notices that a child has missed a problem that the student obviously knows. The teacher stands by the child's desk, taps her finger by the incorrect problem, shakes her head, and walks on to the next desk.	Ethical Unethical	3.6 96.4	3.5 96.5	3.6 96.4

Table 3
Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices in multiple assessment opportunities

Item #	Scenarios for multiple assessment opportunities	Respondents' answers	Total (N = 169), %	Pre-service teachers (N = 114), %	In-service teachers (N = 55), %
21	A teacher assesses student knowledge by using many types of assessments: multiple-choice tests, essays, projects, portfolios.	Ethical Unethical	99.4 0.6	99.1 0.89	100 0
10	A high school social studies teacher bases students' final semester grade on 2 multiple-choice tests.	Ethical Unethical	21.9 78.1	22.8 77.2	20 80
25	A second-grade teacher uses observations as the sole method to assess what students have learned.	Ethical Unethical	15.4 84.6	14.9 85.1	16.4 83.6

Most of the in-service and pre-service teachers rated as unethical a teacher disclosing to parents the achievement scores of a child other than their own (98%, Item 16), and identifying individual students' rankings on exams to others (93%, Item 20).

3.3. Comparisons between in-service and pre-service teachers

We conducted chi-square analyses to compare the responses of in-service teachers and pre-service teachers. With 36 items for comparisons, the Type I error would be inflated. Thus, we applied the Bonferroni correction to our test-wise alpha of .05 to reduce our Type I, or false positive, error rate to .0014 (i.e., $0.05 \div 36$ comparisons). Only on 1 of the

36 items, did in-service teachers and pre-service teachers provide ratings that were significantly different. For item 4 (i.e., a teacher creates learning activities with specific math problems that are included in the annual achievement test), pre-service teachers were significantly more likely to rate the practice as ethical than were in-service teachers ($\chi^2 = 13.80$, $df = 1$, $p < 0.001$). Thus, with one exception, teaching experience did not account for differences in labeling an assessment practice as ethical or unethical.

3.4. Analysis across content categories

To examine patterns in agreements and disagreements, we tabulated the number of items in each

Table 4
Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices in communication about grading

Item #	Scenarios about communication about grading	Respondents' answers	Total (N = 169), %	Pre-service teachers (N = 114), %	In-service teachers (N = 55), %
1	A teacher states how she will grade a task when she assigns it.	Ethical Unethical	98.2 1.8	97.4 2.6	100 0
27	A teacher tells students what materials are important to learn in preparing for a class test.	Ethical Unethical	98.2 1.8	97.4 2.6	100 0
31	A middle school principal directs teachers to give students a written policy that explains how report card grades are calculated in their classes.	Ethical Unethical	97.6 2.4	98.2 1.8	96.4 3.6
29	For the final exam, a teacher always uses a few surprise items about topics that were not on the study guide.	Ethical Unethical	33.7 66.3	35.1 64.9	30.9 69.1

Note: All items in this table were classified as "Do No Harm".

Table 5
Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices in grading practices

Item #	Scenarios about grading practices	Respondents' answers	Total (N = 169), %	Pre-service teachers (N = 114), %	In-service teachers (N = 55), %
24	A teacher lowers grades for late work by one letter grade for each day.	Ethical Unethical	85.8 14.2	86.8 13.2	83.6 16.4
30	A teacher considers student effort when determining grades.	Ethical Unethical	85.2 14.8	88.6 11.4	78.2 21.8
3	For a group project, a teacher bases each student's grade on the group's product and a heavily weighted individual component.	Ethical Unethical	83.4 16.6	83.3 16.7	83.6 16.4
33	A teacher considers a student's growth in assigning grades.	Ethical Unethical	79.3 20.7	78.9 21.1	80.0 20.0
15	To encourage lively discussion in English III, a teacher counts class participation as 30% of the final grade.	Ethical Unethical	74.6 25.4	72.8 27.2	78.2 21.8
32	A teacher weights homework heavily in determining report card grades.	Ethical Unethical	57.4 42.6	57.9 42.1	56.4 43.6
8	A physical education teacher gives a student a zero as a homework grade for not returning a form requiring a parent's signature.	Ethical Unethical	42.6 57.4	42.1 57.9	43.6 56.4
13	As a teacher finalizes grades, she changes one student's course grade from a B+ to an A because tests and papers showed the student had mastered the course objectives even though he had not completed some of his homework assignments.	Ethical Unethical	37.3 62.7	33.3 66.7	45.5 54.5
14	To minimize guessing, a teacher announces she will deduct more points for a wrong answer than for leaving the answer blank.	Ethical Unethical	30.8 69.2	29.8 70.2	32.7 67.3

Table 5 (continued)

Item #	Scenarios about grading practices	Respondents' answers	Total (N = 169), %	Pre-service teachers (N = 114), %	In-service teachers (N = 55), %
19	A middle school history teacher offers extra credit opportunities to all his classes except the advanced class.	Ethical Unethical	26.0 74.0	21.9 78.1	34.5 65.5
28	A teacher uses student peer ratings as 40% of the grade on an oral report.	Ethical Unethical	25.4 74.6	21.1 78.9	34.5 65.6
11	An accounting teacher gives a student an F for the course because the student missed the final exam.	Ethical Unethical	24.9 75.1	24.6 75.4	25.5 74.5
35	A teacher lowers report card grades for disruptive behavior.	Ethical Unethical	15.4 84.6	18.4 81.6	9.1 90.9

Table 6

Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices items in bias

Item #	Scenarios about bias	Respondents' answers	Total (N = 169), %	Pre-service teachers (N = 114), %	In-service teachers (N = 55), %
34	A teacher allows a student with a learning disability in the language arts to use a tape-recorder when the student answers the essay questions on social studies tests.	Ethical Unethical	94.1 5.9	93.9 6.1	94.5 5.5
26	A teacher always knows the identity of the student whose essay test she is grading.	Ethical Unethical	48.5 51.5	42.1 57.9	61.8 38.2
36	To enhance self-esteem, an elementary teacher addresses only students' strengths when writing narrative report cards.	Ethical Unethical	40.8 59.2	39.5 60.5	43.6 56.4
18	A teacher who knows a student had a bad week because of problems at home bumps the student's participation grade up a few points to compensate for his bad score on a quiz.	Ethical Unethical	24.9 75.1	24.6 75.4	25.5 74.5
22	Two teachers teach different sections of the same course. Because of his belief that students' work is rarely perfect, one teacher gives very few grades of "A".	Ethical Unethical	20.1 79.9	18.4 81.6	23.6 76.4

category that displayed high agreement (i.e., 80% or higher) and those that displayed high disagreement (i.e., 50–70%). As shown in Table 8, the categories with high agreement on individual items among respondents included *Communication about Grading* (with 3 of the 4 items reaching at least 80% agreement), *Confidentiality* (2 of the 3 items), and *Multiple Assessment Opportunities* (also 2 of 3). The category of *Grading Practices* had the lowest

agreement among responses with only 4 of the 13 items reaching 80% agreement), followed by *Bias*, with 2 of the 5 items reaching 80% agreement. Overall, 17 of the 36 items fell into the high agreement range (80–100% agreement).

In examining the seven content categories, we also noted items with high levels of disagreement. Table 8 shows that 5 of the 7 categories contained items with high levels of disagreement, *Standardized*

Table 7
Percentage of pre-service and in-service teachers indicating the ethicality of assessment practices in confidentiality

Item #	Scenarios about confidentiality	Respondents' answers	Total (N = 169), %	Pre-service teachers (N = 114), %	In-service teachers (N = 55), %
9	A teacher discloses to the parents of a student their child's score on an intelligence test.	Ethical	75.7	76.3	74.5
		Unethical	24.3	23.7	25.5
20	To motivate students to perform better, a science teacher always announces that he is passing out scored tests to students in order of points earned, from the top score to the bottom score.	Ethical	6.5	5.3	9.1
		Unethical	93.5	94.7	90.9
16	To calm the fears of distraught parents, a teacher compares their child's achievement scores with the results of the student's cousin who is also in the class.	Ethical	2.4	1.8	3.8
		Unethical	97.6	98.2	96.4

Note: All items in this table were classified as "Do No Harm".

Table 8
Categories ranked by percent of items showing high agreement (80%+) and high disagreement (40–70%)

Category	Items showing high agreement	Items showing high disagreement
Communication about grading	3 (75%)	1 (25%)
Multiple assessment opportunities	2 (67%)	0
Confidentiality	2 (67%)	0
Standardized test preparation	3 (50%)	1 (17%)
Standardized test administration	1 (50%)	1 (50%)
Bias	2 (40%)	2 (40%)
Grading practices	4 (31%)	4 (31%)

Note: Rows do not sum to 100% because some items did not meet the criterion of 80–100% for high agreement or the criterion of 50–70% for high disagreement.

Test Administration (1 of 2 items), *Bias* (2 of 5 items), *Grading Practices* (4 of 13 items), *Communication about Grading* (1 of 4 items), and *Standardized Test Preparation* (1 of 6 items). Overall, 9 of the 36 items fell into the high disagreement range (40–70% agreement), with 10 items in the medium agreement range (71–79% agreement).

3.5. Agreement analysis on Do No Harm and Score Pollution items

In examining items related to the two principles, five out of seven *Do No Harm* items had high agreement 71.4%, and only one had high disagreement (14.3%). In addition, *Do No Harm* items were all in the top two agreement categories. Only 12 out of 28 *Score Pollution* items had high agreement (43%) and 8 had high disagreement (29%), suggest-

ing more variation in responses for *Score Pollution* items.

3.6. Summary of results

In summarizing results for all respondents, *Communication about Grading* had the highest levels of agreement, with 75% of the items in the high agreement range and one item in the high disagreement range. The categories *Confidentiality* and *Multiple Assessment Opportunities* also had high levels of agreement, with a majority of items (67%) in the high agreement range and no items falling in the high disagreement range. The categories with the greatest split between respondents were *Standardized Test Administration*, *Bias*, and *Grading Practices*, with equal numbers of items in the both high agreement and the high disagreement ranges

and no more than 50% of the items in the high agreement range. In addition, *Do No Harm* items reached higher levels of agreement than *Score Pollution* items. Finally, in-service and pre-service teachers responded similarly to all but one item.

4. Discussion

Each year the media report the instance of an educator who violates some norm of acceptable assessment practices. The assessment literature provides some guidance for teachers in terms of ethical and unethical practices in standardized testing; however, few resources in assessment directly address the ethics of classroom assessment practices. This initial effort to develop scenarios that address principles of ethical assessment practice indicates that pre-service and in-service teachers at these universities had strong agreement on less than half of the scenarios presented in this study. These findings suggest that assessment is a realm without professional consensus. Thus, courses related to instruction and assessment should address the needs of educators in these areas. Strike (1990) suggests that specific moral concepts are unlikely to be addressed in pre-service teacher education curriculum. He recommends explicit instruction in ethical concepts, grounded in the activities that are vital to the teaching profession. Thus, conversations about principles such as *Do No Harm* and *Avoid Score Pollution* as they relate to daily assessment practices in schools could be productive. In addition, research has shown that support for teachers embedded in everyday details of their ongoing activities is an effective approach for influencing beliefs and practices of teachers (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997). These findings suggest that discussions in schools around ethical principles and classroom assessment could also be useful.

One area that teachers experienced disagreement was that of grading practices, the area that created the firestorm in a Midwestern community. Only 4 of the 13 items reached high agreement, and two of these items contrasted with the recommendations of the assessment community. Most assessment experts agree that grades should reflect only the extent to which students have mastered the goals of instruction (e.g., Brookhart, 2004; Smith et al., 2001), yet a strong majority of respondents (85%) rated as ethical considering effort or late work when calculating grades (Item 30). Similarly, a majority of teachers rated as ethical the lowering of students'

grades for late work (Item 24, 86%). Again, this is an issue of score pollution because the grade begins to convey information about behavior (i.e., tardiness in completing work) rather than providing information about a student's understanding of the subject matter. These findings suggest that the issue of score pollution in classroom grading should be highlighted in future work.

The area of score pollution is important to continue to address in future research and in the classroom because of the complexity and nuance involved. For example, to avoid score pollution, a grade on a group project should involve an individual component and not just one grade for the entire group (Woolfolk, 2003). If the group component is weighted significantly higher than the individual component, score pollution can result because grades then have a stronger probability of not reflecting the individual student's mastery of the learning goals. Similarly, teachers often include dimensions, such as neatness or class participation, in their grading schemes that are not directly related to the mastery of the learning goals. If the percentage devoted to such factors has minimal impact on the overall grade, such practices could be ethical. However, if such factors are weighted heavily enough to change a grade, they may result in score pollution. Score pollution is one issue where theory meets reality in the classroom. The college incident where professors were fired for refusal to allocate 60% of their grade to effort is a case in point.

While we also see the guiding principle of *Do No Harm* as a valuable one, we recognize it is not unproblematic. What constitutes harm is itself often a matter of judgment within context. Even within the realm of assessment, there are a variety of harms that could be done to students. There is the potential educational harm done as the result of assessments that fail to accurately measure the knowledge or skills that they claim to measure. There is also the potential emotional harm done to students in the form of anxiety or other stress that high-stakes assessments often bring about. There is also the potential for harm of the teacher–student relationship. Teacher–student trust can be damaged by assessments that the student perceives as unfair or unfounded.

At times, *Do No Harm* emerges in practice as a choice between harms. A teacher may have to choose between a high-stakes exam that yields important data about student performance and the

emotional stress that such exams cause in students (and teachers). We see this situation (the choice between harms) as potentially more “normal” than the avoidance of harm altogether.

Clearly, the *Do No Harm* formulation of assessment ethics is not, in and of itself, without complexities and problems. We believe, however, that it serves as a useful beginning to approaching the issues of our research.

To further expand the point of complexity, educators’ responses probably reflect the extent to which respondents are bound by normative rules in their own teaching situations. For example, some school districts might purchase test preparation materials for use in the schools, and thus teachers might consider using such materials as ethical. When presented with the scenario “A teacher uses *Scoring High on the MAT*, a commercially available publication with the same format and skills as the Metropolitan Achievement Test (but not the same items), in preparation for state testing,” teachers in such school districts might indicate the practice is ethical. In other settings, however, educators who use test preparation materials will find such practices considered to be unethical and their actions open to sanction (Natrona County School District, 2001; North Carolina Department of Public Instruction, 2000; South Carolina Department of Education, 2005).

One of the limitations of the current research method was the requirement that respondents endorse a practice as either ethical or unethical based on limited information. The use of only two options implies no possibility of a gray area and requires limited, noncontextual responses. The contexts established in the scenarios require interpretation, and educators might arrive at differing decisions due to nuances in their interpretations of the context. Some educators might consider the nuances that are implicit in a scenario and see a given assessment practice as being ethical in some situations and unethical in others, creating a gray area for the ethicality of a practice. The scenarios, however, create an opening for discussion of the ethics of practice. In future research, the use of open-ended responses and/or a range of factual variations for each type of scenario could yield interesting comparisons.

The designation of items as ethical or unethical in Section 4 was based on a survey of the assessment literature and the opinion of leaders in the assessment field. Given that some of these issues do

involve nuances, the findings from the current study provide a picture of the degree of agreement with these insights by this group of pre-service and in-service teachers. This study has been an effort to initiate a dialogue and discussion of these important issues. As the frequent newspaper anecdotes attest, a more consistent sense of ethical assessment practices among teachers, as well as the general public, is needed. Implementing ethical assessment in the classroom is not simply a matter of advocating “best practice.” A number of teachers have paid a significant price for violating what has been perceived, rightly or wrongly, as ethical assessment practice in their communities.

References

- Airasian, P. (2005). *Assessment in the classroom: A concise approach* (2nd ed.). Boston, MA: McGraw-Hill Company.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990). *Standards for teacher competence in educational assessment of students*. Retrieved November 2, 2004 from <http://www.unl.edu/buros/article3.html>.
- Baumgart, N. (1996). Codes of ethics on assessment practice: Confronting a quandary or inculcating ethical character? Overview of two sample documents. *Assessment in education: Principles, policies & practice*, 3.
- Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers’ developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, 13, 259–278.
- Bright, E. (1992). Teachers’ views of ethical standardized test use. Paper presented at the *annual meeting of the National Council on Measurement in Education*. San Francisco, CA.
- Brookhart, S. (2004). *Grading*. Upper Saddle River, NJ: Pearson.
- Carroll, D. (2002). Piper teacher who resigned in plagiarism dispute is a hit with media. *The Kansas City Star*. Retrieved August 3, 2004 from <http://www.kansascity.com/mld/kansascity/news/local/2634836.htm?1c>.
- Click, C. (2004). Benedict fires two professors bucking effort-based grading. *The State*. Retrieved December 15, 2004 from http://www.thestate.com/mld/thestate/news/breaking_news/9447674.htm.
- Colnerud, G. (1997). Ethical conflicts in teaching. *Teaching and Teacher Education*, 13, 627–635.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. Washington, DC: The Falmer Press.
- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham: Open University Press.

- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20, 2–7.
- Hostetler, K. D. (1997). *Ethical judgment in teaching*. Boston: Allyn and Bacon.
- Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Educational administrators' and teachers' knowledge of classroom assessment. *Journal of School Leadership*, 3, 510–521.
- JCSEE. (2003). *The student evaluation standards*. Arlen Gullikson, Chair. Thousand Oaks, CA: Corwin.
- Joint Advisory Committee. (1993). *The principles for fair student assessment practices for education in Canada*. Edmonton, Alberta. Retrieved July 15, 2004 from <http://www.acposb.on.ca/fair.pdf>.
- Jonsen, A., Siegler, M., & Winslade, W. (1998). *Clinical ethics*. New York: McGraw Hill.
- Kant, I. (1785). Foundation for metaphysics of morals (Thomas Kingsmill Abbott, Trans.). In W. Kaufmann, & F. Baird (Eds.), *From Plato to Nietzsche* (pp. 834–877). Englewood Cliffs, NJ: Prentice-Hall.
- Kant, I. (1790). *Critique of judgment* (J.H. Bernard, Trans.). New York: Hafner Publishing Company, Inc. 1951.
- Kher-Durlabhju, N., & Lacina-Gifford, L. (1992). Quest for test success: Preservice teachers' views of "high stakes" tests. Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN.
- Kilian, L. J. (1992). A school district perspective on appropriate test preparation practices: A reaction to Popham's proposals. *Educational Measurement: Issues and Practice*, 11(4), 13–16.
- Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice*, 8(1), 14–22.
- Messick, S. (1993). Validity. In R. Linn (Ed.), *Educational measurement* (2nd ed., pp. 13–103). Washington: American Council on Education.
- Moore, W.P. (1993). Preparation of students for testing: Teacher differentiation of appropriate and inappropriate practices. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.
- Munson, R. (2000). *Intervention and reflection: Basic issues in medical ethics*. Belmont, CA: Wadsworth Publishing.
- National Association for the Education of Young Children (2005). *Code of ethical conduct and statement of commitment*. Retrieved December 27, 2005, from <http://www.naeyc.org/about/positions/pdf/PSETH05.PDF>.
- Natrona County School District (2001). *Board policy # 6302: Test ethics*. Casper, Wyoming: Retrieved March 21, 2006, from <http://mail1.ncsd.k12.wy.us>.
- Noddings, N. (1984). *Caring: A feminine approach to ethics and moral education*. Berkeley: The University of California Press.
- North Carolina Department of Public Instruction (2000). *Testing code of ethics*. Retrieved March 21, 2006, from <http://www.ncpublicschools.org/docs/accountability/testing/policies/testcode080100.pdf>.
- Nussbaum, M. (1986). *The fragility of goodness*. Cambridge: Cambridge University Press.
- Nussbaum, M. (1990). *Love's knowledge*. New York: Oxford University Press.
- Payne, D. A. (2003). *Applied educational assessment* (2nd ed.). Belmont, CA: Wadsworth Publishing.
- Plake, B., & Impara, J. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment*. San Diego: Academic Press.
- Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice*, 10(4), 12–15.
- Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders*. Needham, MA: Allyn & Bacon.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Smith, J. K., Smith, L. F., & DeLisi, R. (2001). *Natural classroom assessment: Designing seamless instruction and assessment*. Thousand Oaks, CA: Corwin Press.
- South Carolina Department of Education. (2005). *PACT administration manual: Spring 2005*. Columbia, SC: Author.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23–27.
- Strike, K. (1990). Teaching ethics to teachers: What the curriculum should be about. *Teaching and Teacher Education*, 6, 47–53.
- Strike, K., & Soltis, J. (1998). *The ethics of teaching* (3rd ed.). New York: Teachers College Press.
- Taylor, K., & Nolen, S. (2005). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, NJ: Pearson Education, Inc.
- Thorndike, R., Cunningham, G., Thorndike, R., & Hagen, E. (1991). *Measurement and evaluation in psychology and education* (5th ed.). New York: McMillan.
- Woolfolk, A. (2003). *Educational psychology: Developing learners* (4th ed.). Upper Saddle River, NJ: Merrill Prentice-Hall.